DOCUMENT RESUME

ED 185 971 IR 008 296

AUTHOR Hardy, Nancy F.
TITLE Maximizing the Effectiveness of Online Searching--A Training and Education Model.
PUB DATE 79
NOTE 10p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Bases; *Models; *Online Systems; *Search Strategies; *Training Methods; Worksheets

ABSTRACT

Three aspects of training and education for online searching of bibliographic and databases are addressed: (1) the present status--what is being taught, (2) a proposed model--what needs to be taught, and (3) the implementation of the model--who should teach what. The first section shows which organizations (library and information schools, information retrieval services vendors, database producers, and other online user groups) are engaging in which education and training activities: question negotiation, search strategy, evaluation of results, information retrieval system mechanics, database-specific information, and database loading characteristics by retrieval system. Basic competencies, search strategy development, and evaluation of results are the three major components of the model presented in the second section. A sample worksheet is included. Finally, who should have the responsibility for teaching different aspects of online searching is briefly considered in the conclusion. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185971

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MAXIMIZING THE EFFECTIVENESS OF ONLINE SEARCHING--A TRAINING AND EDUCATION MODEL

by

Nancy F. Hardy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Nancy F. Hardy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR008296

MAXIMIZING THE EFFECTIVENESS OF ONLINE SEARCHING – A TRAINING AND EDUCATION MODEL

Nancy F. Hardy, Engineering Index, Inc., U.S.A.

1 INTRODUCTION

This paper, "Maximizing The Effectiveness of Online Searching -- A Training and Education Model," will address three aspects of training and education for online searching of bibliographic databases. These are 1) the present status -- that which is being taught, 2) a proposed model -- that which needs to be taught, and 3) the implementation of the model -- who should teach what. There have been at least two papers describing the training activities of database producers in the last year. One of these, written by Betty Unruh (Data Courier) was published in the Proceedings of the 2d International Online Meeting held in London in December 1978. (Ref. 1) The other was a presentation made at one of the 1979 American Library Association/Machine Assisted Reference Section programs which were held in Dallas during June of this year. (Ref. 2) At the June RCD/MARS panel on training for online searching, educational activities of database producers, information retrieval system vendors, library and information schools, and online user groups were described. It is from presentations such as these, as well as active participation in various education and training groups, that this aggregate picture of training activities for online searching has been accumulated.

2 PRESENT STATUS OF TRAINING ACTIVITIES

There are four major types of organizations involved in education and training activities for online searching. These are 1) library and information schools, 2) information retrieval service vendors, 3) database producers, and 4) a mixed group which is composed of online user groups, regional bibliographic centers (e.g., the Pacific Bibliographic Center), commercial organizations (e.g., Cuadra Associates), and networking organizations (e.g., AMIGOS in the U. S. and UTLAS in Canada). These organizations are currently teaching in the areas of:

1. Question negotiation (a pre-search function)
2. Search strategy
3. Evaluation of search results
4. Information retrieval system mechanics
5. Database specific information
6. Database loading characteristics by retrieval system

In Figure 1 is a display indicating which organizations are engaging in which education and training activities.

ACTIVITY	LIBRARY/INFORMATION SCHOOLS	INFORMATION RETRIEVAL SYSTEMS	DATABASE PRODUCERS	OTHERS
Question Negotiation	Y	N	N	Y
Search Strategy	Y	Y	Y	Y
Evaluating Results	Y	*	N	N
System Mechanics	Y	Y	Y	Y
Database Information	*	*	Y	N
Database Loading	N	N	Y	N

Y = Yes
N = No
* = Some

Figure 1. Education and Training Activities by Organization

The asterisks in the columns indicate minimal participation in a given training activity. For example, in the area of providing database specific information. Figure 1 indicates that library and information schools and information retrieval systems provide minimal training activity.

3 PROPOSED MODEL FOR TRAINING ACTIVITIES

The training model presented here recommends some of the concepts, skills, and techniques which need to be acquired in order to maximize the effectiveness of online searching. The three major components are Basic Competencies. Search Strategy Development, and Evaluation of Results. (Figure 2) Although the individual components of the model will be discussed, the model is not intended to be exhaustive, in that there is still much to be learned about the attributes of a component online searcher.

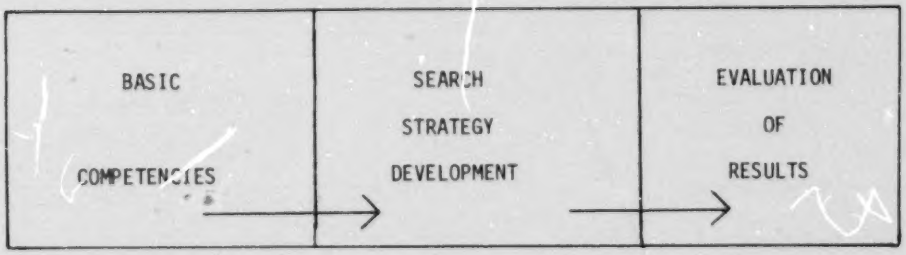


Figure 2. Model of Training Needed For Effective Searching

3.1 Basic Competencies

The basic competencies needed for effective online searching include thorough knowledge of the information retrieval system mechanics,

database specific information, and how a given database is loaded on a given retrieval system. (Figure 3.)

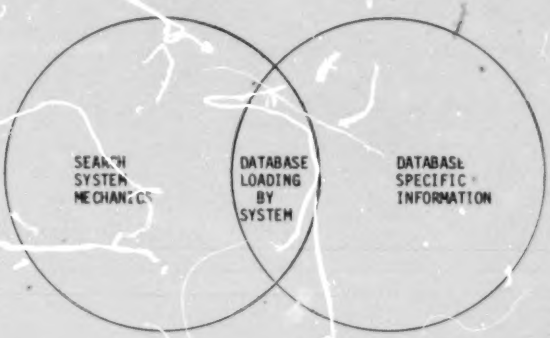


Figure 3. The "BASICS" Necessary For Effective Online Searching

Thorough knowledge of system mechanics or protocols is a continuing educational process which enables the searcher to keep up with new developments in the information retrieval search systems. Examples of new developments in system mechanics include:

1. SUPERSELECT on DIALOG
2. LIMIT on CAN/OLE
3. LINK in ORBIT
4. PRINT/set number/document type command (enabling the user to order hard copy or microfiche) in IRS/RECON
5. New log on procedures using TELENET in the U. S. Department of Energy RECON system.

Another basic need for effective online searching is database specific information. This means the editorial, selection, coverage, and indexing policies of the database which one is attempting to search. Such details as the handling of geographic indicators vary considerably from database to database, and if one needs this type of geographic information, it is imperative to understand how geographic names are treated in the database in question. This includes where the information is placed in the record, as well as the form of the name used, particularly in the case of political entities. For example, should one look for geographic place names in the corporate source field, in an indexing field, or perhaps in the abstract field. Will the format of the name be West Germany or the Federal Republic of Germany; Peoples Republic of China, Mainland China, or China; Taiwan, Formosa, or Republic of China? Other kinds of database specific information include such things as the language of the title of the article (English, vernacular, or both), whether or not the language of the original paper is indicated if an English language abstract is provided, and whether or not authority files are used for corporate names.

A third fundamental piece of knowledge which is necessary for effective online searching is "file loading." This is an almost entirely neglected area in the education and training of online searchers, for reasons to be discussed later in this paper. It is <u>critically</u> important

for the searcher to know how a given database is loaded - the programming constraints applied to the given database - on the specific information retrieval system being used. I will give examples from the COMPENDEX database, since that is the one with which I am most familiar. Can one search the abstract, present on a COMPENDEX tape, in the CAN/OLE system? Can one search the Main Heading - Subheading combination in the format as assigned by the Engineering Index editors in the DIALOG system using only the inverted file? Can one search for non-English titles in the ORBIT system? Or can one print the entire bibliographic citation when it is very lengthy in the IRS/RECON system? These and many other questions can only be answered by knowledge of how a given database is loaded by an information retrieval system. File loading characteristics truly impact the effectiveness of searching any particular database online.

3.2 Search Strategy Development

In addition to the basic competencies previously identified, another area in online training which has not been adequately covered is search strategy development. Three components which have been identified are 1) understanding indexing (specifically the use of vocabularies and thesauri), 2) copying with natural language or text searching, and 3) identifying "hidden" access points.

Numerous books and articles have been written on indexing -- its concepts, principles, the construction of vocabularies and thesauri, and the application of the indexing terms. Nevertheless, there seems to be widespread misunderstanding and lack of knowledge concerning what indexing is and that which indexing can and cannot do. Although forecasters predict that the full text of documents and articles will be stored online, that is still in the future for most bibliographic databases. Because of inherent difficulties in using natural language searching (to be discussed later), indexing will be with us for some time to come, and it can provide a highly-effective means of accessing documents which contain pertinent information. It is not the purpose of this paper to explore the advantages and disadvantages of indexing a database for online retrieval. Rather, it is suggested that a better understanding of indexing itself, as well as the indexing policies of various database producers (including those vocabularies and thesauri which may be used to index a database) is essential for effective online searching.

These points may seem more than obvious to most searchers; however, it is not at all uncommon for an individual to assume that a certain vocabulary or thesaurus is used to index a given database, without actual knowledge of whether or not such is truly the case. A very common misunderstanding is that the Engineering Index database, COMPENDEX, is indexed using a thesaurus. This is certainly not the case because the indexing language used is a rather simple list of subject headings; it is certainly not a thesaurus. In addition, it is commonly believed that the Thesaurus of Engineering and Scientific Terms and/or the Engineering Index Thesaurus are the indexing vocabularies for COMPENDEX. This is also incorrect. The vocabulary used to index COMPENDEX is named SHE (Subject Headings for Engineering). It is easy to understand why these misconceptions exist; however, it is the responsibility of the searcher to identify which vocabulary is used to index a given database when index terms are used in developing the search strategy.

Coping with natural language searching (or full text searching) is considerably more complex than may be imagined. To begin with there are

several variations of full text searching, including what has been called "true" text searching, proximity searching, and adjacency searching. Some, but not all, of these text searching capabilities are available on most information retrieval systems. The inherent problems in searching text, referred to earlier, have been explored at length in another paper. (Ref. 3) To review some of these problems, it may be said that there are at least five major problem areas in text searching. They are:

1. Syntax
2. Orthography
3. Semantics
4. Grammar
5. Symbolism

Searching problems such as word order, American versus English spelling, ambiguity of words, usage of words, multiple forms of the same word, symbols, formulas, punctuation, etc. are not so formidable when the searcher has a minimal understanding of some of these linguistic and semiotic principles. Solutions can be found and are available in most information retrieval systems to overcome many of these difficulties. However, at the present time, there seems to be scant information transmitted to online searchers in these particular areas.

There are several access points in bibliographic database records which are either little-used, ineffectively used, or not used at all. In this paper they are called "hidden" access points. Three of these are journal/publication codes (CODEN, ISSN, ISBN, etc.), corporate names, and the bibliographic citation itself. Examples of utilizing "hidden" access points follow.

1. To limit a search in COMPENDEX to AGARD (Advisory Group for Aerospace Research and Development) publications, include the following set of CODEN:

   AAAPBK
   AGCHAV
   NAGLBS
   AGARAR
   AGAGAS

   or

2. To increase the number of hits in COMPENDEX when geographic indicators are desired, judiciously include the locations indicated with corporate names:

PLACE NAME	CAN/OLE POSTINGS	
	TITLE INDEX	ORGANIZATION INDEX
ALBERTA	203	634
MANITOBA	46	224
BRITISH COLUMBIA	144	309

3. To identify an entire conference on the use of computers in cardiology, in COMPENDEX include a truncated character string search for "comput" and "cardiol" in the citation field.

COMPENDEX Worksheet

TITLE:

SEARCH ELEMENTS: CONCEPT A CONCEPT B CONCEPT C

SHE TERMS:

NATURAL LANGUAGE:
- TITLE
- AUTHOR'S AFFILIATION
- BIBLIOGRAPHIC CITATION
- ABSTRACT

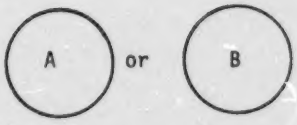
CODEN:

CAL CODES:

AUTHORS:

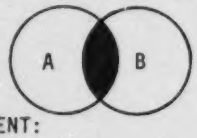
BOOLEAN LOGIC OPERATIONS:

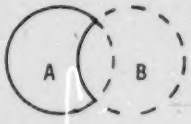
1. "OR" TERMS IN EACH COLUMN.



2. "AND" or "AND NOT" THE COLLECTION OF TERMS IN COLUMN A WITH THE COLLECTION OF TERMS IN ALL OTHER COLUMNS.

A and B





A and not B

SEARCH STATEMENT:

ALTERNATE SEARCH STATEMENT:

Figure 4. Search Strategy Development Worksheet

3.3 Evaluation of Results

The evaluation of retrieval results is frequently taught using measures of recall and precision. Additional questions which can and should be asked in order to evaluate results are:

1. Were the appropriate databases used? Too many? Too few?
2. Was the retrieval appropriate for each database used, both in quantity and quality?

Thorough knowledge of specific databases enables the searcher to answer these and other evaluative questions. It is highly suspect, for example, to retrieve only four or five citations (all of them in Russian) from COMPENDEX when searching for documents about solar energy collectors. One should quite reasonably expect a fairly large retrieval on this topic from the COMPENDEX database.

4 CONCLUSION

The proposed model for training and educational activities reiterates a long-standing dilemma -- Who should teach which components of the model? This question has been debated for at least three years, and no simple resolution of the problem has been proposed. The ASIS (American Society for Information Science) Online Educators User Group, the Committee on the Education and Training of Search Analysts of the American Library Association and other groups have addressed the problem without firm recommendations. Therefore, the following is a proposed distribution of the education responsibilities by activity and type of organization.

Principles of computer-based information systems which are essential to the understanding of online searching should be taught in library and information schools. Techniques for evaluating search results, question negotiation, search strategy principles, principles and practices in indexing, beginning system mechanics, and some elementary semiotics could also be taught in library and information schools.

Strategy, advanced system mechanics, basic database information and evaluation could also be taught by information retrieval services. Database producers should be responsible for providing specific and detailed information about their files. Additionally, it is strongly felt by the author that the teaching of the characteristics of file loading is the database producer's responsibility.

Comments on this model and its implementation are encouraged by the author.

REFERENCES CITED

1. Unruh, Betty. User Education - The Channel for Communication. (IN 2nd International Online Information Meeting, London, 5-7 December, 1978. Paper B7, pp. 65-74. Oxford, Learned Information (Europe) Ltd., n.d.)

2. Hardy, Nancy F. Education and Training Activities of Database Producers. Presentation at "Options for Online Training," 1979 American Library Association Annual Conference, Dallas, 24 June, 1979.

Unpublished paper.

3. Hardy, Nancy F. Natural Language Is Not So "Natural." ONLINE '79. Atlanta, 7 November, 1979.

10